humans amidst automation: competition, learning, and uncertainty

Roy Dong (University of California, Berkeley)

Lillian J. Ratliff (University of Washington)

digital competence

↓

digital usage

↓

digital transformation


4:06PM
35:19.7
135

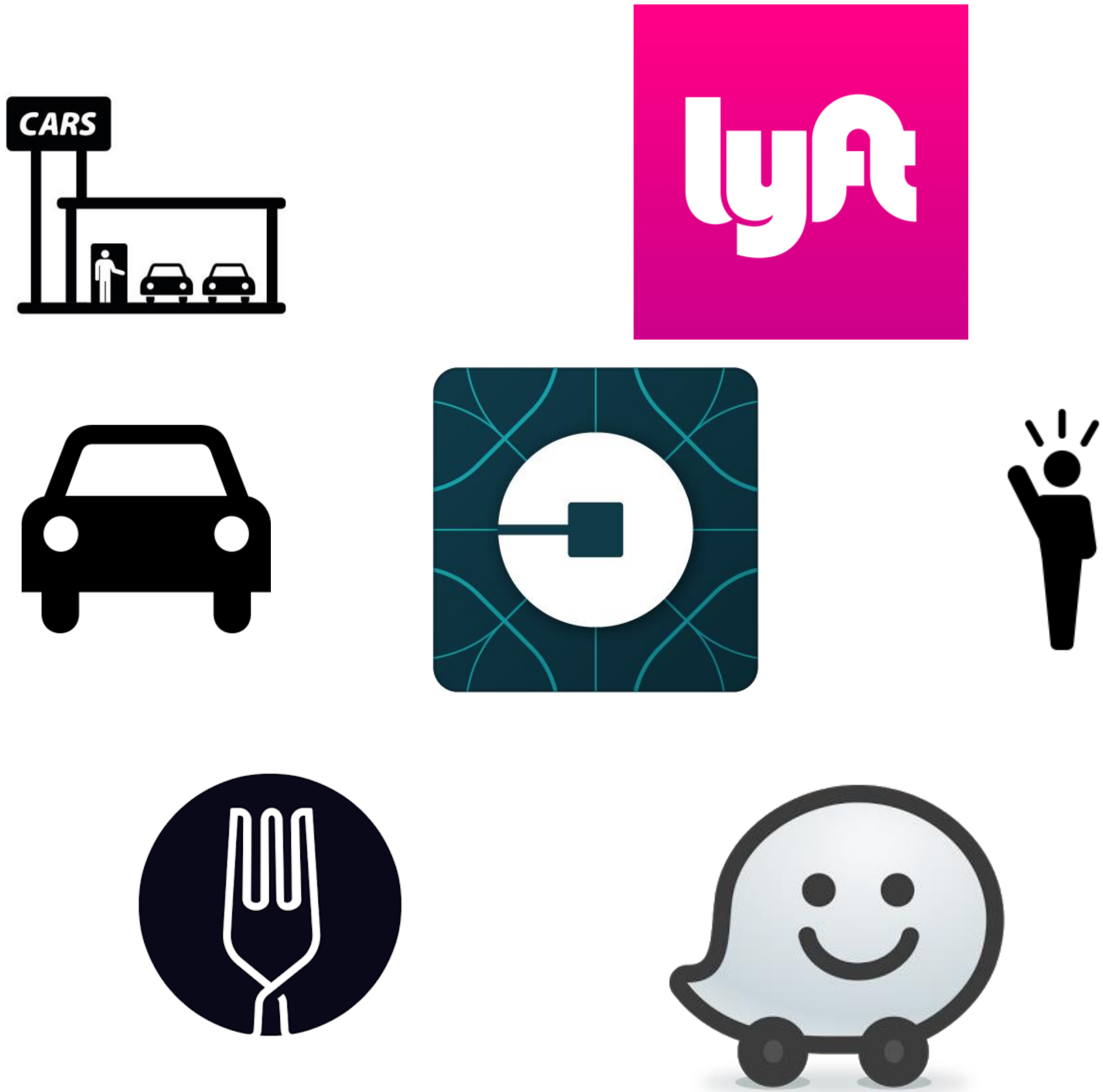
Total Time Sitting Before 6PM*
6
5hr 35min
5hr 45min
5hr 47min
US
fitbit

TWELVE 12OZ B
STONE

The **digital transformation** for IoT data:

- Data will enter into these systems in a *closed-loop* fashion through analytics, controllers, and incentives.
- Users will have incentive to obfuscate and strategically manipulate their data.
- Companies will have to compete for consumers, and will use data to improve their competitive edge.

n-sided markets



n-sided markets

- Looking forward, we wish to understand the effects of **competition**, **asymmetric information**, and **learning** in these n-sided markets.

- To start:
  - With one firm: how do we **learn** the preferences and behaviors of users?
  - With one firm: how do we **close the loop** on this learning process?
  - With multiple firms: what is the effect of **competition**?

outline

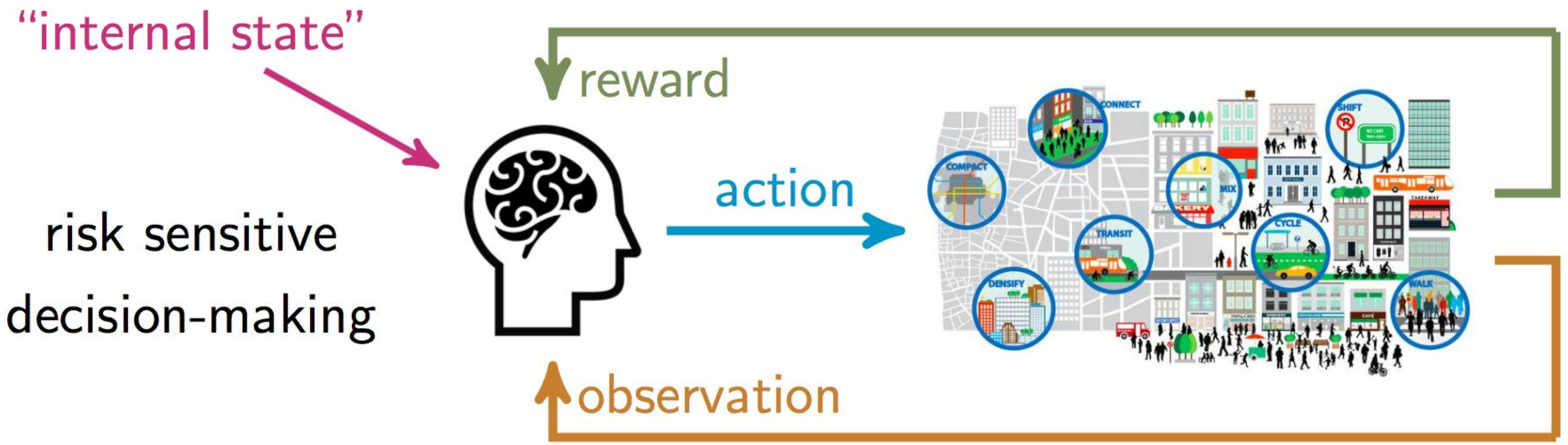
- learning
  - inverse reinforcement learning with risk-sensitive agents
- learning and control
  - multi-armed bandit approaches for issuing incentives when preferences and dynamics are unknown
- competition
  - equilibria of data markets

outline

- learning
  - inverse reinforcement learning with risk-sensitive agents
- learning and control
  - multi-armed bandit approaches for issuing incentives when preferences and dynamics are unknown
- competition
  - equilibria of data markets

Learning: Inverse Risk-Sensitive RL

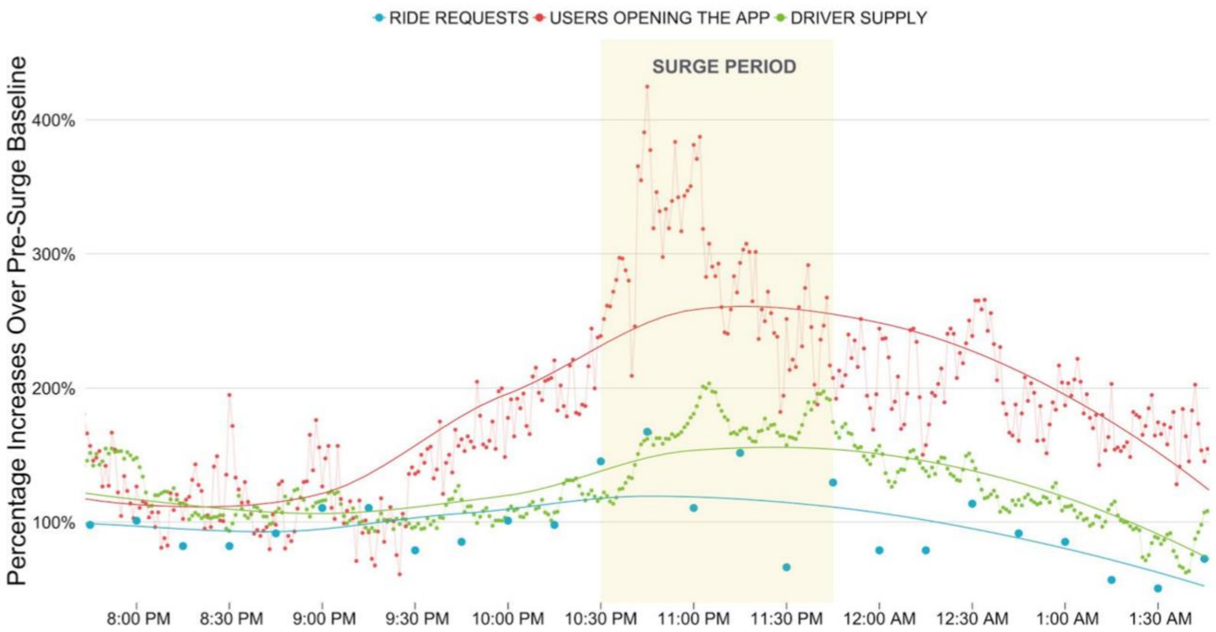
Can we learn plausible models of human behavior and preferences, with theoretical foundations, by drawing on "smart" infrastructure data?



- Humans tend to treat losses and gains differently & make decisions based on reference points and distortions of event probabilities.
- challenge: rational, utility maximization models tend not to capture these effects

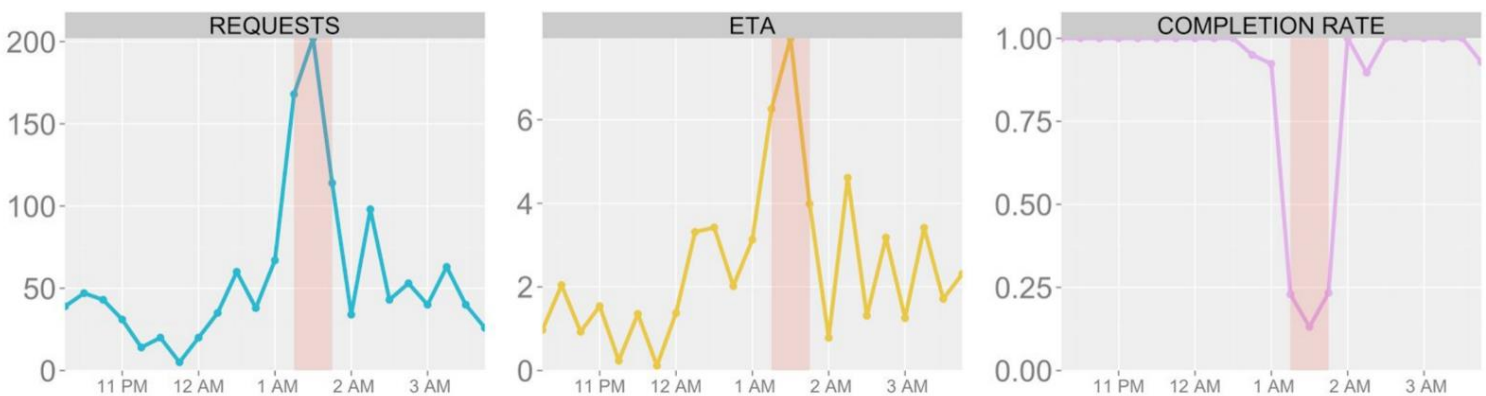
Goal: leverage fine grained user choice data to develop (real-time) algorithms for learning and designing incentives in closed loop

Uber Case Study—Losses Loom Larger than Gains



drivers: marginal gain

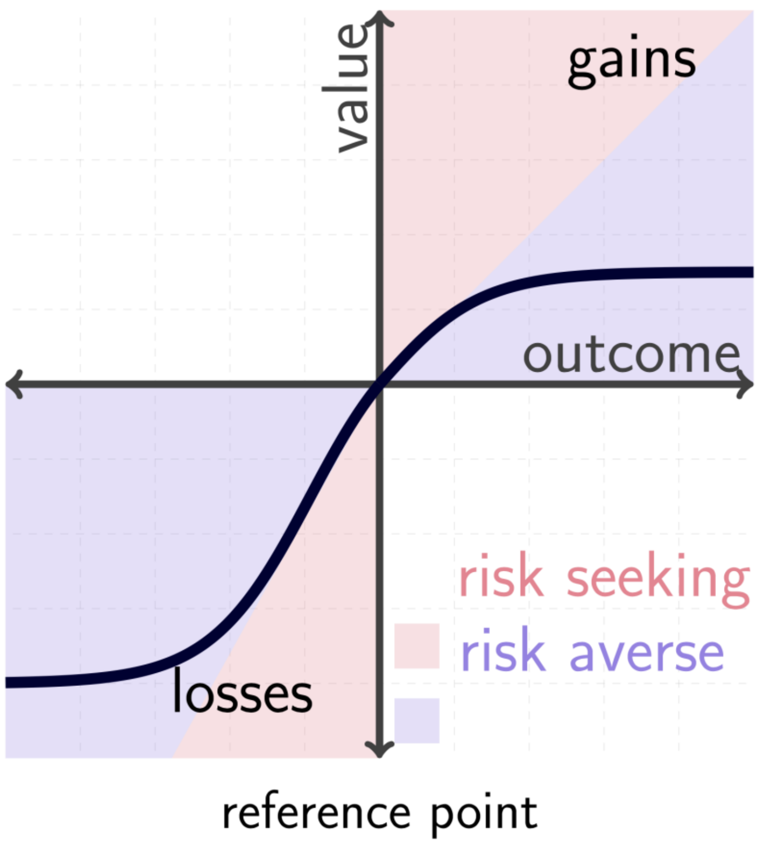
passengers: major loss



NYC Sold-Out Concert, March 21, 2015 (credit: J. Hall *et al.*, 2015)

Salient Features: Loss Aversion and Risk-Sensitivity

- reference points—e.g.,
  - status quo
  - recent expectations about future
- outcomes compared to reference points
  - more preferable outcomes are gains, otherwise a loss
  - losses tend to loom larger than gains
- risk-attitudes impacted by reference points
  - more risk-averse on gains (concave)
  - more risk-seeking on losses (convex)

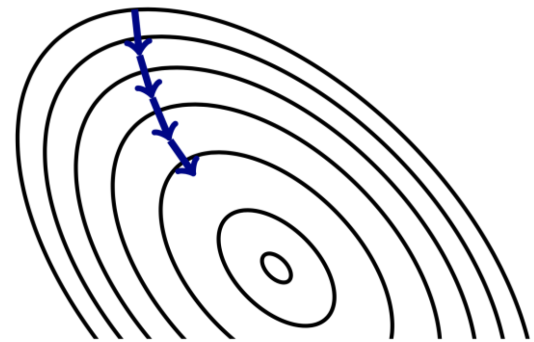


$$u(y) = \begin{cases} k_+(y - r_0)^{l_+}, & y \geq r_0 \\ -k_-(r_0 - y)^{l_-}, & y < r_0 \end{cases}$$

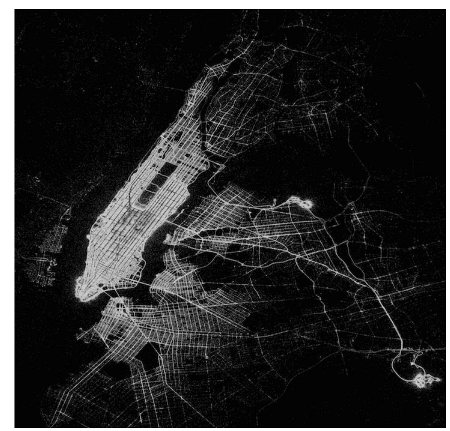
Inverse Risk-Sensitive RL

- **Model**: To capture these salient features, we couple **behavioral models** w/ **coherent risk metrics** in a MDP model.
- e.g., **short-fall risk**: compare to outside option

$$\mathscr{V}(Y) = \sup\{m \in \mathbb{R} \mid \mathbb{E}[u(Y-m)] \geq u_0\}$$



- **Learning**: (gradient-based inverse RL) we learn the parameters of the value function, learning process, and acceptance level
- **Convergence**: assuming a Q-learning process, we derive contraction map argument for Q and its derivative w.r.t. parameters
- **Application**: classical gridworld, Uber data (passenger's view), and NY taxi data (in progress)



outline

- learning
  - inverse reinforcement learning with risk-sensitive agents
- learning and control
  - multi-armed bandit approaches for issuing incentives when preferences and dynamics are unknown
- competition
  - equilibria of data markets

Learning & Control: Preference Dependent Incentives

- Preferences evolve over time & depend on incentives offered
- **Multi-Armed Bandits (MAB):** assume preferences evolve according to a Markov process with a different transition kernel associated with each arm

Challenges:

- Assuming one type of agent, the problem is still challenging because the arms are dependent.
- With multiple types, the problem is combinatorial.
- Exploration may lead to volatility in incentive which can cause agents to opt out.

$$\theta_i^- \xrightarrow{\mu^j} \theta_i^+$$

$$\theta_i^- \xrightarrow{\mu^j} \theta_i^+$$

$$\theta_i^- \xrightarrow{\mu^j} \theta_i^+$$

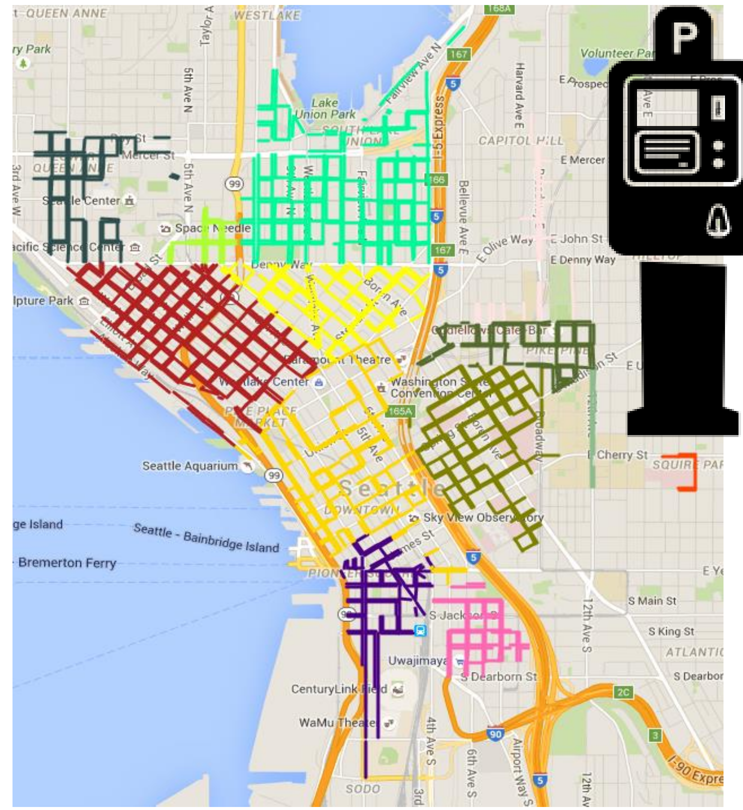
Approach:

- Exploit dependencies to reduce complexity.
- Introduce risk-metrics (MV, CV@R, & other coherent risk measures) metricize volatility
- Provide usual regret bounds for UCB-type algorithms (depends on duration an arm is selected and size of type space: O(log(n)) single type and O(M^{3}Nlog(n)), M=#users, N=#resources)

Living Lab: Smart Parking & Targeted Ads/ Incentives

SDOT currently:

- sets prices based on data from one sample per year of occupancy levels
- targets incentives and performance-based pricing in pre-defined, static neighborhoods



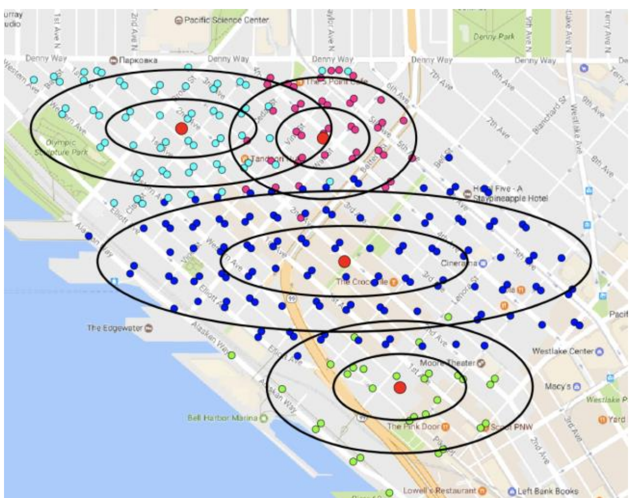
This leads to poor performance and **unintended consequences** such as reduced business district vitality and increased congestion

SDOT
Seattle Department of Transportation

Living Lab: Smart Parking & Targeted Ads/ Incentives

Create experiments in Seattle to target ads and incentives to locations with similar characteristics (e.g., demand and business type).

Gaussian Mixture Model to identify locations with similar demand



Target **ads** (e.g., locations of typically available parking, off-street options, etc.) & **incentives** (e.g., bus passes, coupons to local businesses)

Working with SDOT **marketing** team & **Business Improvement Areas**

motivation

- Deep learning can achieve great performance in control…

motivation

- …even when there are known sensor failures…

motivation

- … and there's recent theoretical work towards proving their optimality.
  - Neural networks can express a very, very, very large class of functions. [Raghu, Poole, Kleinberg, Ganguli, Sohl-Dickstein 2017] [Zhang, Bengio, Hardt, Recht, Vinyals 2017]
  - All local optima are global optima under some positive homogeneity assumptions. [Haeffele, Vidal 2015]
  - In deep residual networks, under certain assumptions, when the network is deep and wide enough: every stationary point is a global optimum, there are no local minima, and no saddle points. [Bartlett, Evans, Long]

motivation

- However, sensor failures are often **unknown** online.
- Suppose we have a set of **experts** $\{e_1, e_2, \ldots, e_N\}$, each trained for a different failure mode.
- How can we choose our expert **online** to minimize our **regret**?

Joint work with Eric Mazumdar and Vicenç Rúbies Royo.

problem formulation

- Our model of the world is an partially observed Markov decision process (MDP), denoted $(\mathcal{S}, \mathcal{A}, \mathcal{Y}, P, O, R, \mu_0)$.

- Each expert maps observations $\mathcal{Y}$ to actions $\mathcal{A}$.

- If we listened to expert e_i for all time, then we'd get average reward $\bar{R}_i = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} R_i(t)$.

problem formulation

- The average reward of expert e_i:

$$\overline{R}_i = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} R_i(t)$$

- Since we don't know the best expert, we will use our expert's suggestions to pick actions $a(t)$ and get rewards $R(t)$.

- We define **regret** at time t:

$$\max_i t\overline{R}_i - \sum_{s=0}^{t-1} r(s)$$

problem formulation

$$\max_i t\bar{R}_i - \sum_{s=0}^{t-1} r(s)$$

- The **contribution** of our work is that we have the experts control the **same system.**
  - If we listen to expert e_i at time t and expert e_j at time $t+1$, the reward $r(t+1)$ will be influenced by e_i's performance.
  - There is lots of coupling between experts in this formulation!
  - **Intuition:** We commit to an expert for long enough for these transient effects to die out.

theoretical results

- At iteration n listen to expert $e(n)$ for T_n time steps.

- A new regret decomposition identity:

$$\mathbb{E}[r(n)] \leq \sum_{e \neq e^*} \mathbb{E}[T_e(n)] \left[\Delta_e + \frac{C_e}{T_0(1-\alpha_e)}\right] + \frac{C_*}{1-\alpha_*} \sum_{k=0}^{n-1} \frac{1}{T_k}$$

theoretical results

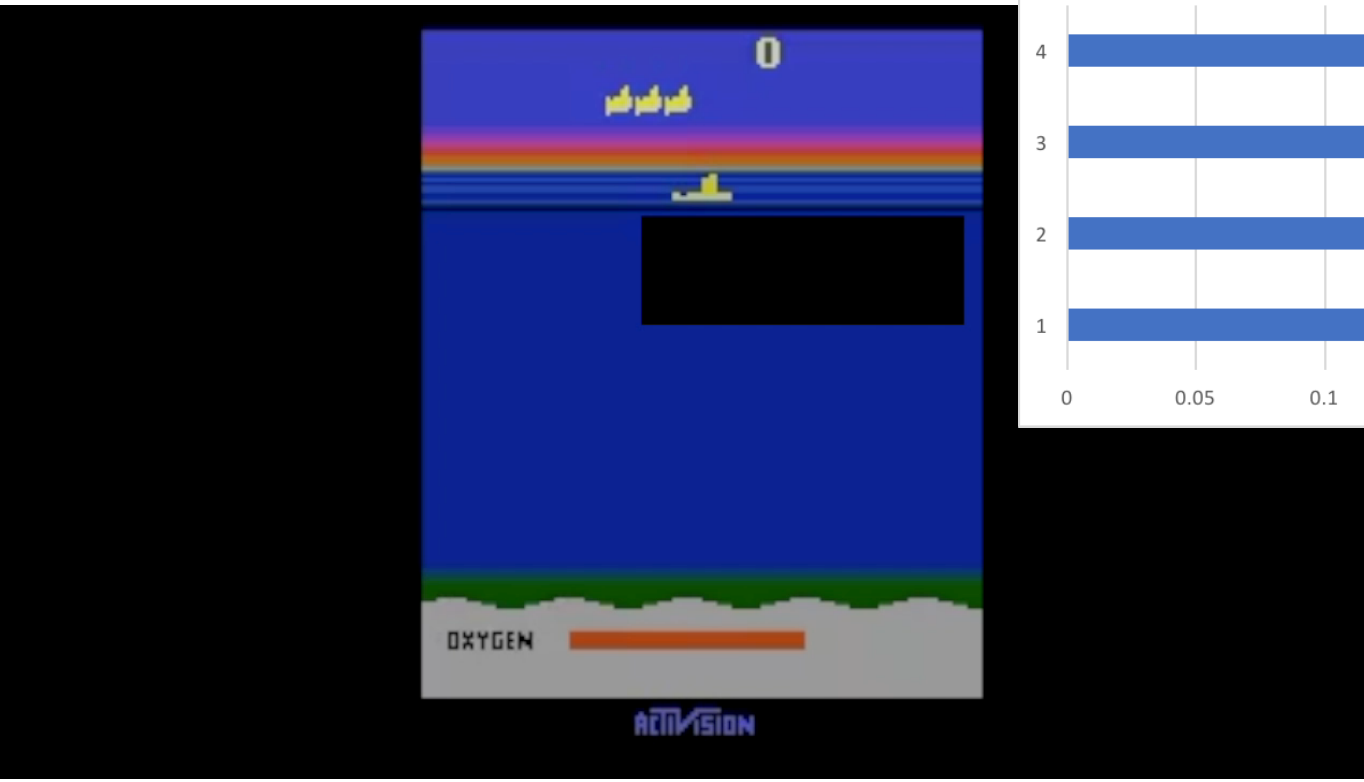
- At iteration n listen to expert $e(n)$ for T_n time steps.

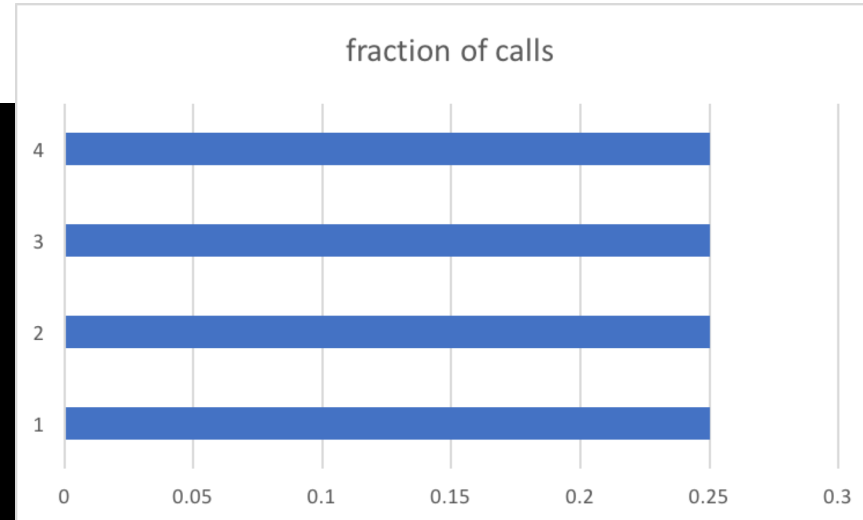
- Using the upper confidence bound algorithm, with $T_n = \lfloor T_0 + cn \rfloor$:

$$\mathbb{E}[r(n)] \leq \sum_{e \neq e^*} \left[\left(\frac{32\log(n)}{\left(\Delta_e - \frac{2K_e}{T_n}\right)^2} + 1 + \frac{\pi^2}{3} \right) \left(\Delta_e + \frac{K_e}{T_0} \right) \right] + \frac{K_*}{c} \log\left(\frac{T_0 + cn - c}{T_0} \right)$$

simulation results

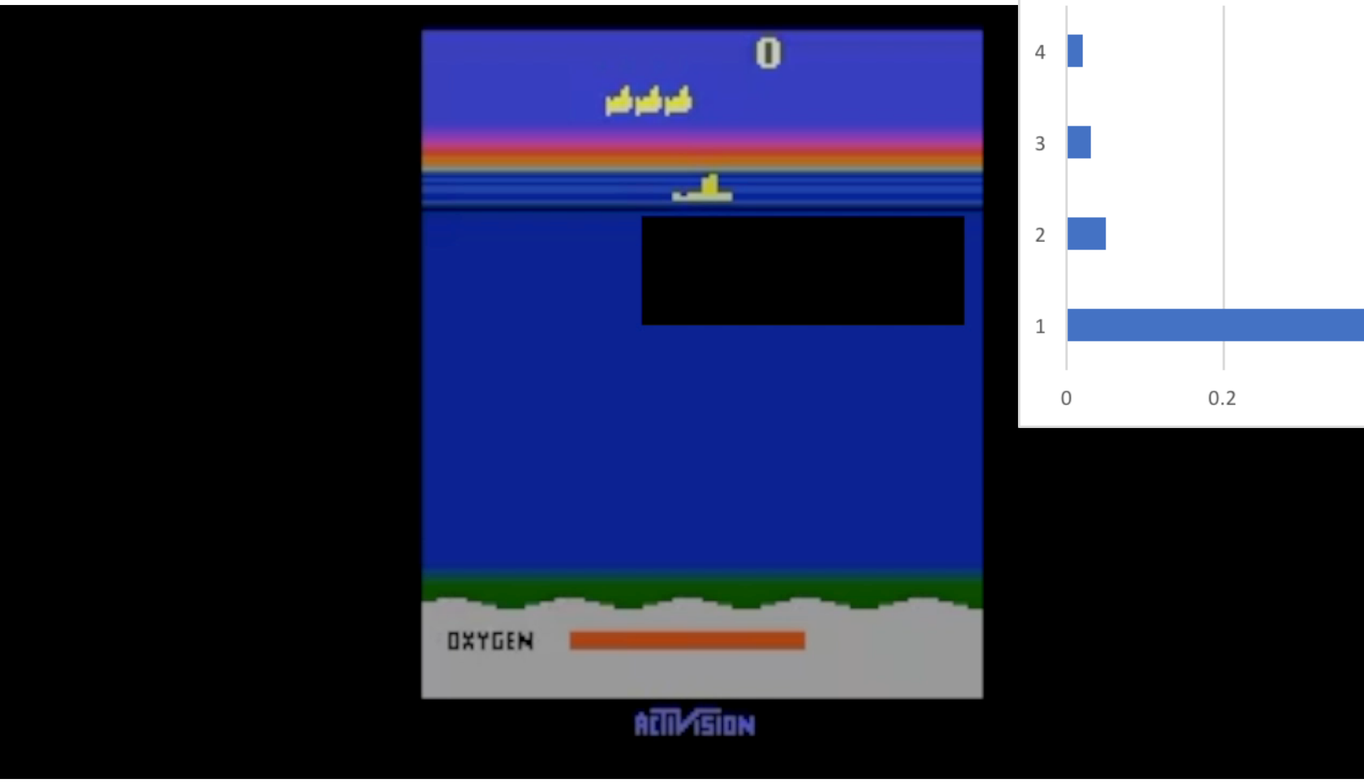
- Initially…

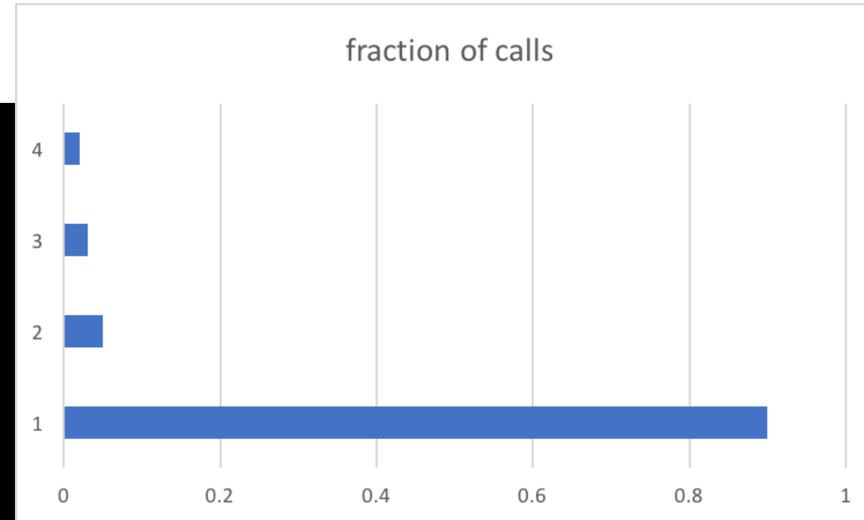




simulation results

- After 1250 iterations…

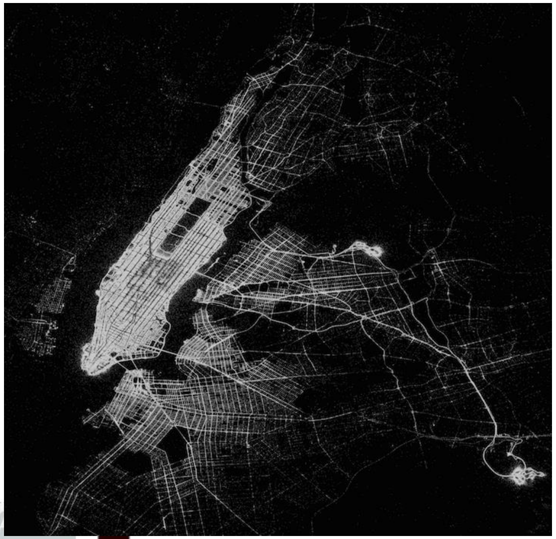




From Experts to Incentives

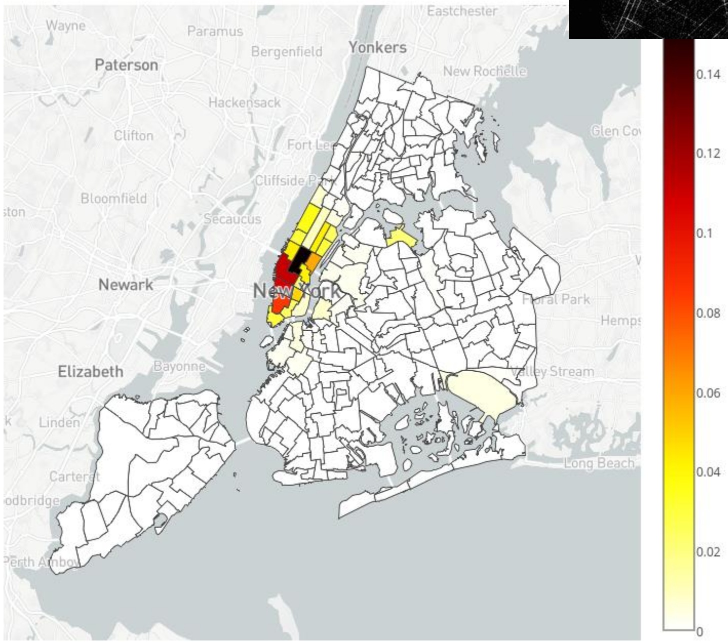
- Extension of Experts MAB to incentivizing risk-sensitive agent:

$$\underbrace{R_e(s,a,s')}_{\text{reward for expert}} \longrightarrow \underbrace{f(\pi_j(s,a))\ (\text{or } f_j(s,a,s'))}_{\text{principal observed reward for arm } j}$$



- NY Taxi data from 2010-2014 (~30k drivers)
- Derive a MDP model of taxi drivers via inverse (risk-sensitive) RL
- Learned MDP model is used to simulate drivers and we design reward functions for the principal
- e.g., incentives to visit areas with high-demand

New York Taxi Data: Portion of Rides per Area

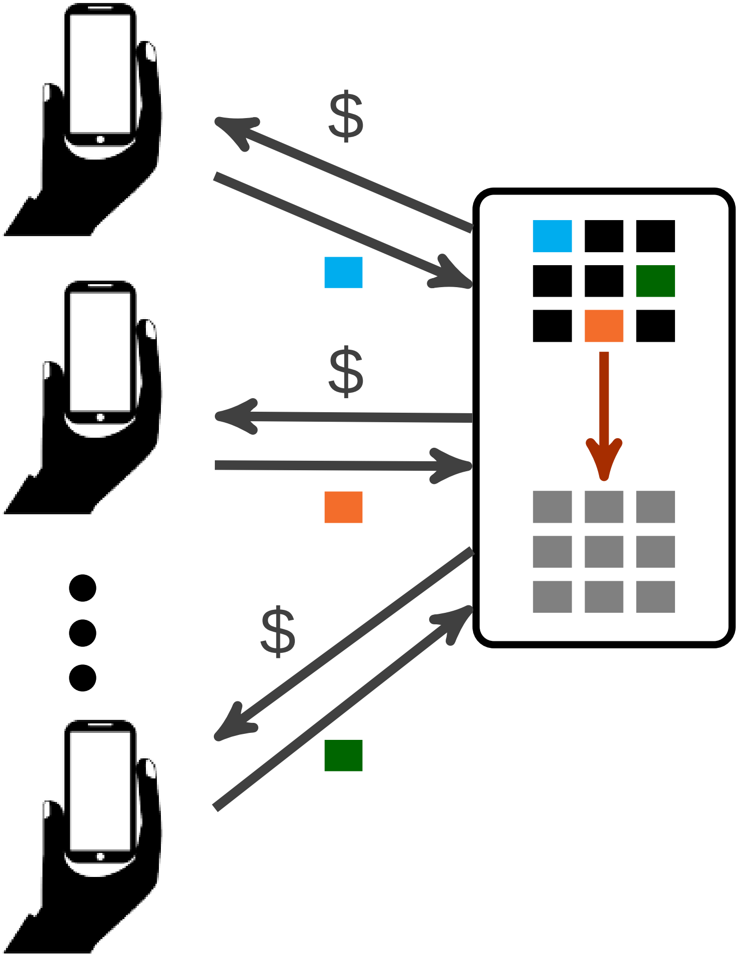


outline

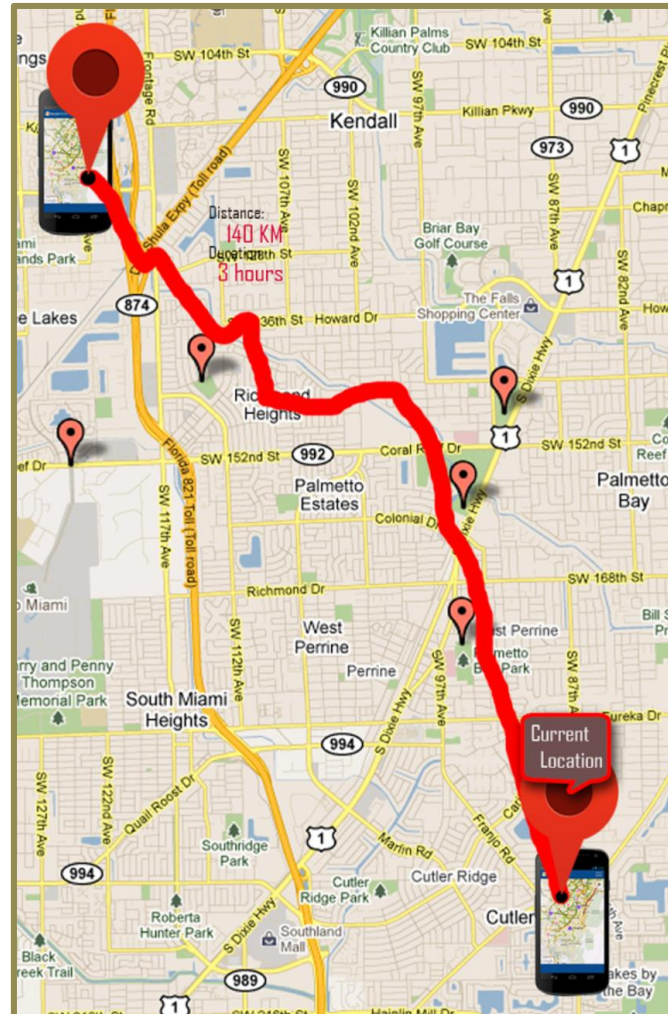
- learning
  - inverse reinforcement learning with risk-sensitive agents
- learning and control
  - multi-armed bandit approaches for issuing incentives when preferences and dynamics are unknown
- competition
  - equilibria of data markets

data markets and services

- In our recent work, we model the **data market**:
  - users exchange their data in exchange for services and incentives
  - data buyers balance their statistical estimation goals with the cost of providing incentives
  - multiple data buyers may compete



Joint work with Tyler Westenbroek.


Distance:
140 KM
Duration:
3 hours
Current Location
Kendall
Killian Palms Country Club
SW 104th St
Killian Pkwy
Briar Bay Golf Course
The Falls Shopping Center
Richmond Heights
Palmetto Estates
Palmetto Bay
SW 152nd St
SW 168th St
Richmond Dr
West Perrine
Perrine
South Miami Heights
Southridge Park
Cutler Ridge
Cutler Ridge Park
Roberta Hunter Park
Southland Mall
SW 216th St
Hainlin Mill Dr

model

- Strategic data sources are **effort-averse**.
  - They need to be incentivized to provide a certain quality of data.

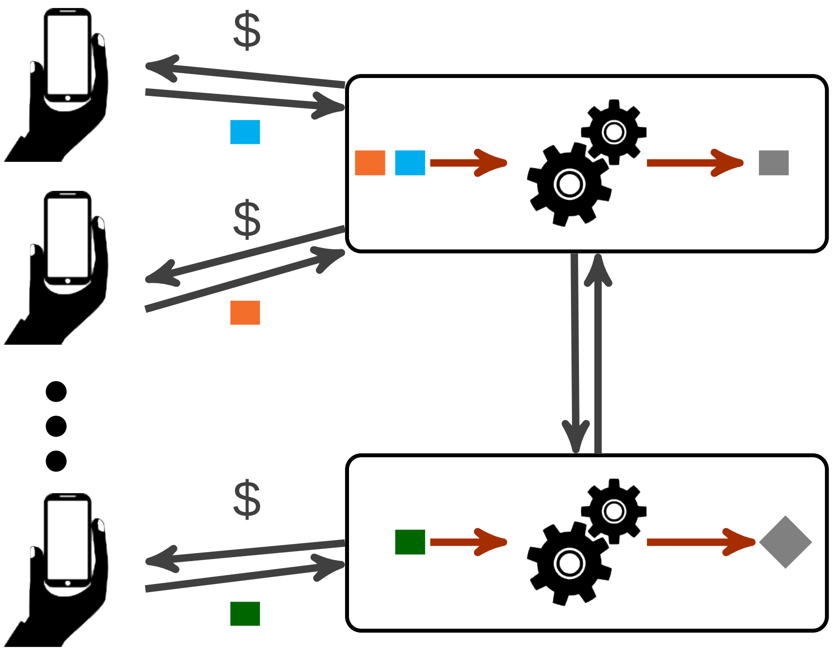
- Data is **non-rivalrous**.

- Multiple data buyers want data sources to exert **sufficient** effort, but don't want to **personally** pay for it.
  - The **total** incentives must justify the effort exerted.
  - The **individual** incentive from data buyer j must incentivize sharing data with j.

data markets summary

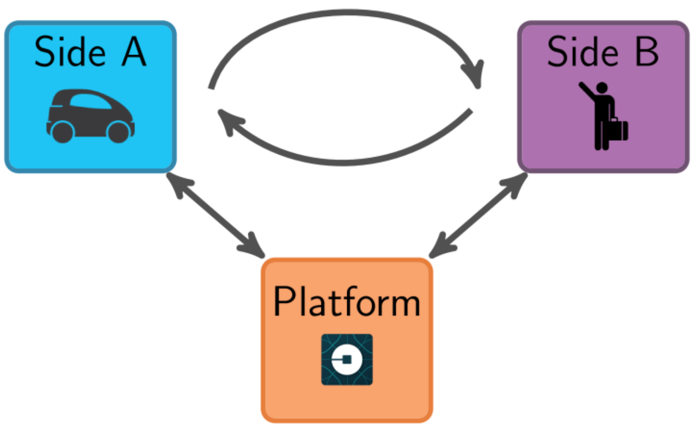
- We developed an **game-theoretic** model to account for strategic data sources, reproducibility of data, quality of statistical estimation, and competition between buyers.

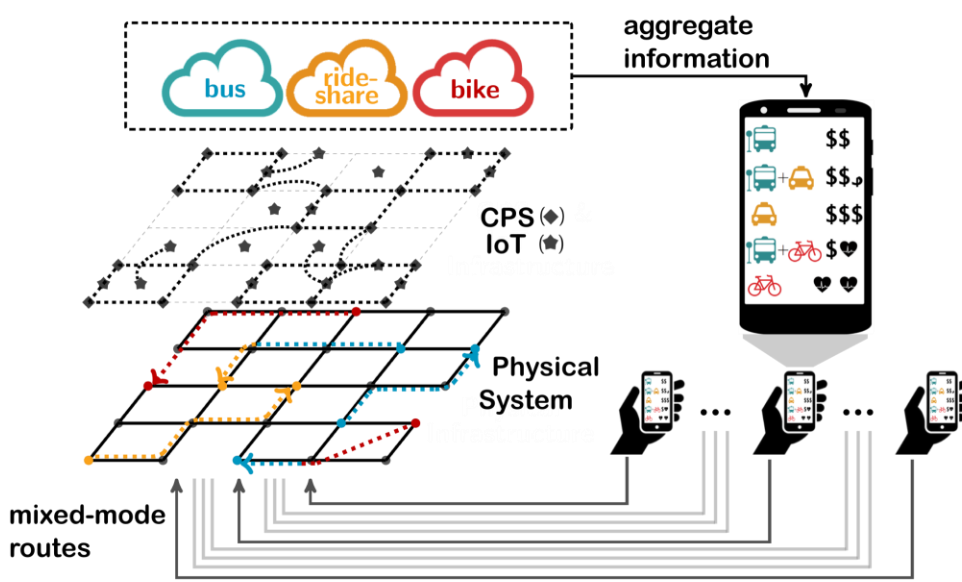
- In contrast to the single-buyer case, when we introduce **competition** between firms, many of the **desiderata** of our incentive schemes are not preserved.



Multi-sided Markets: Matching & Learning via Bandits

- Platform based firms aim to match supply to demand
- Given unknown supply and demand characteristics, we are combining machine learning approaches for segregating (clustering) each side of the market and matching clusters
- e.g., drivers and passengers with similar ratings is a heuristic for matching, but how does this extend when there are multiple objectives such as distance, hours worked, other in-place incentives, etc.





outline

- learning
  - inverse reinforcement learning with risk-sensitive agents
- learning and control
  - multi-armed bandit approaches for issuing incentives when preferences and dynamics are unknown
- competition
  - equilibria of data markets

The Digital Transformation & New Directions

- New research directions on new market structures formed by pervasive, disruptive technologies that serve as the impetus for the **digital transformation**
- These new directions grew out of the methodologies and approaches created by FORCES
  - how **learning** can be done when the data is generated by strategic human agents operating in unstructured, uncertain environments
  - how **competition** interacts with control and estimation of cyber-physical systems
  - how agents and markets respond to **uncertainty**

digital competence

↓

digital usage

↓

digital transformation